# Impact de la qualité des données sur les résultats obtenus avec le modèle

**Cinquième article**
Attribution des cas humains français de salmonellose aux principales sources alimentaires selon le type de données de surveillance

De façon attendue, l'information a priori informative a un impact non négligeable sur les résultats d'attribution, en particulier lorsque cette information est choisie de façon arbitraire. Parmi les paramétrisations évaluées, la proposition basée sur les types microbiologiques spécifiques à certaines espèces animales a permis d'obtenir des résultats convergents, et cohérents avec la méthode déterministe et les facteurs de risques connus dans la littérature. De plus, les résultats produits quant à l'importance relative des sources, sont significatifs. Cette approche a donc été utilisée pour évaluer l'impact de la qualité des données sur les résultats d'attribution.

En effet, comme présenté au chapitre II, le système de surveillance français permet de disposer parallèlement de données de surveillance active et de données de surveillance passive concernant la contamination des sources animales. La différence entre ces deux types de données concerne (i) la représentativité, structurelle pour les données de surveillance active, inconnue pour les données de surveillance passive, (ii) l'indicateur, prévalences pour les données actives, proportions pour les données de surveillance passive, et enfin (iii) le nombre de sources prises en comptes, 5 pour le jeu de données actives, 9 pour le jeu de données passives. L'impact des caractéristiques (ii) et (iii) est évalué individuellement à l'aide du modèle simple précédemment utilisé (Cf chapitre IV). Enfin, les résultats obtenus avec le modèle bayésien ont été comparés pour le jeu de données comprenant les données de surveillance active d'une part et, celui contenant les données de surveillance passive d'autre part.

Le projet d'article présenté est en cours de rédaction pour soumission à Foodborne Pathogens and Disease. Les références à l'article précédent sont notées de la façon suivante : (David, informative information).

# Attribution of the French human Salmonellosis cases to the main food-sources according to the type of surveillance data

J.M. David*[1,2,3,4], D. Guillemot[2,3,4], N. Bemrah[5], A. Brisabois[6], M. Denis[7], FX. Weill[8], N. Jourdan[9], P. Sanders[1], L. Watier[2,3,4]

1 French Food Safety Agency, BP 90203 Fougères, F-35302, France

2 INSERM, U 657, Paris, F-75015, France

3 Institut Pasteur, Pharmacoépidémiologie et Maladies Infectieuses, Paris, F-75015, France

4 Univ. Versailles Saint Quentin, Faculté de Médecine Paris Ile de France Ouest, F-78035, France

5 French Food Safety Agency, Maisons-Alfort, F-94701, France

6 French Food Safety Agency, Maisons-Alfort, F-94700, France

7 French Food Safety Agency, BP 53, Ploufragan, F-22440, France

8 Institut Pasteur, CNR des _Salmonella_, Laboratoire des Bactéries Pathogènes Entériques, Paris, F-75015, France

9 Institut de Veille Sanitaire, Saint-Maurice, F-94415, France

Corresponding author: Julie David, French Food Safety Agency, Institut Pasteur, BP 90203 35302 Fougères, France. Tel: +33 2 99 94 78 78; Fax: +33 2 99 94 78 99. Email: j.david@afssa.fr

# 1 Introduction

*Salmonella* is the first recorded bacterial cause of foodborne gastroenteritis in France, as well as the first cause of relative hospitalizations and deaths (InVS 2004). It is a major public health concern and as such gave rise to a nationwide intervention at the farm level targeting Enteritidis and Typhimurium serotypes in the breeding flocks of laying hens and broilers from 1999 on. The *Gallus gallus* plan has been successful at mitigating the number of human cases (Poirier, Watier et al. 2008), as other similar plans conducted in other European countries (Edel 1994; Wegener, Hald et al. 2003), showing the efficiency of intervening at an early point in the food chain. Though, the challenge remains as *Salmonella* is still a first plan public health issue (Sofos 2008). Attributing human cases to the potential food-sources is a key step to identify, prioritize and assess the impact of interventions from the farm level on (Batz, Doyle et al. 2005). The microbial subtyping attribution approach, which allows attributing human cases at the reservoir point, has already been applied successfully for such purposes in Denmark. It relies on the model published in 2004 by Hald *et al*, and which requires representative data on microbial type repartition both in human and in food animals (Hald, Vose et al. 2004; Mullner, Jones et al. 2009).

To answer such issues, the French integrated surveillance system for *Salmonella* has been developed in the European and national regulatory frame (David, Danan et al. submitted). It covers the whole food-chain from the breeding farms to the human cases and is thus well adapted to the microbial subtyping attribution approach.

Many actors are involved in this surveillance system, which allows disposing of different and complementary expertises (David, Danan et al. submitted). The data on the human cases are collected by the National Reference Centre (NRC), and the National Public Health Institute (InVS) centralizes the data on outbreak related cases which are submitted to mandatory declaration. For what concerns the data on the sources contamination, several databases are available. For this work, we selected the data collected by the Food Directory

of the French Agriculture Ministry (DGAL) in the frame of the institutional surveillance. We also used data collected in the frame of a generalist passive surveillance network (*Salmonella* Network, SN) of the French Food Safety Agency (AFSSA). The first dataset gathers representative prevalence data per microbiological type for the main channels, ideal for attribution. The passive dataset of SN, allows disposing of data on supplementary food-animal sources, which can be important as one of the assumptions of the attribution model is that all cases are linked exclusively to the included sources (Pires, Evers et al. 2009). We thus applied the microbial subtyping approach with both types of data, as to assess the opportunity to use the passive data in such an approach as a complement for active data, and studied the impact of the differences between the two kind of animal data on the model results. For this purpose, two different datasets were constituted, gathering the human data on Salmonellosis cases and either active surveillance data or passive surveillance data on the food-animal sources.

The impact of some of the data characteristics on the attribution results has first been studied, using a simple model (David, informative information), which allows having a more immediate insight on the interaction between the data and the model (Zwietering 2008). The general comparison of the attribution results obtained according to the surveillance type, was made using the Bayesian model described by Hald *et al* (2004), with an adapted informative prior, as proposed in a former paper (David, informative information).

# 2 Materials and methods

## 2.1 Data

### 2.1.1 Human cases

The human cases data were retrieved from NRC and InVS databases for the year 2005, which was the last year for which complete information on the strains was available (serotype and subtype). The cases correspond to confirmed non-Typhi Salmonellosis cases which strains, or analysis reports, have been sent on a voluntary basis to the NRC by public and private microbiology laboratories, together with epidemiological information.

As the model is designed to attribute domestic sporadic cases, i.e. cases which are not related to an outbreak and have been infected on the national metropolitan territory, we gathered information on recent travels and outbreaks. Information regarding travelling abroad was available in the CNR database, the outbreak-related cases were identified by confronting the InVS outbreak database and the CNR database. The identified travel and outbreak related case were excluded from the dataset as well as the cases registered in the outseas departments and territories. For each outbreak, one case is kept in the dataset to represent it in the model.

### 2.1.2 Food-animal sources

Among the available datasets on animal food sources (David, Danan et al. submitted), we selected databases retrieving data at the national level, issued either from routine surveillance or large scale, long-term studies (at least one year). The data on the animal sources should have been collected in 2005 to match the human data, be representative and encompass at least the main channels (layers, broilers, pigs, turkeys, cattle). As the microbial subtyping approach attributes at the reservoir-level, we focused on data collected either at the farm level or at the abattoir level, as early as possible in the slaughter line.

## Active surveillance

In the frame of the European regulation, and to prepare the enforcement of the zoonoses directive, one year long national prevalence studies have been conducted from 2005 to 2007 in the main animal channels either at the farm (layers, broilers, turkeys) or at the slaughterhouse level (pigs). For each food-animal species, a representative sample of flocks or carcasses has been tested for *Salmonella*, 519 layer flocks, 371 broiler flocks, 331 turkey flocks and 1 166 pig carcasses. All isolates were serotyped and the estimated prevalences were adjusted for the flock size for layers, broilers and turkeys. Data on cattle come from the national surveillance plan of antimicrobial resistance in indicator and zoonotic bacteria in cattle, a national annual monitoring plan aiming at evaluating in this channel the prevalence of resistant bacteria, among which *Salmonella*.

These data, thanks to the sampling design of the studies they are issued of, are representative of the French prevalence level at the production stage. Though, the results for broilers, turkeys and pigs are not for 2005 but for 2006 (broilers) and 2007 (turkeys and pigs). Indeed, no representative active surveillance studies were available for 2005 for those channels at the measurement points focused.


## Passive surveillance

SN centralizes strains and analysis reports on non human strains of *Salmonella,* sent on a voluntary basis by a network of private and public veterinary laboratories (David, Danan et al. submitted). We extracted the 2005 serotyping data for the 5 main channels plus sheep, ducks, other poultry and sea products. The "sea products" category gathers data on fish, shellfish and seafood, the "other poultry" category gathers data on geese, guinea fowls, pheasants, quails and pigeons. We only included strains collected at the farm level, on production animals and their environment and at the abattoir level on carcasses. Strains relative to breeding animals, to diseased animals or which origin was unclear were excluded.

For cattle, an overrepresentation of pathogenic strains has been identified at the farm level. As a consequence, for this channel, we only included the strains collected at the abattoir level.

## 2.1.3 Subtyping

To optimize the attribution, Enteritidis and Typhimurium strains which represent more than 30% of the human cases each, have been further subtyped. The subtyping is based on antimicrobial resistance profiles (Berge, Atwill et al. 2003). For the human cases, we used the data available for 92 strains out of 3 536 for Typhimurium and on 102 out of 3 138 for Enteritidis. For the food-animal sources, the profiles realized by SN on the strains after exclusion of duplicates has been used for the passive dataset and for the active dataset, all strains of Enteritidis and Typhimurium collected in the frame of the baseline studies have been further analyzed for antimicrobial resistance especially for this study.

Antibiotic susceptibility was determined by disk diffusion on Mueller-Hinton agar according to the guidelines of the Antibiogram Committee of the French Society for Microbiology. All the routine tested antimicrobials common to the human and animal databases have been included: amoxicillin, chloramphenicol, ceftazidim, gentamicin, kanamycin, nalidixic acid, streptomycin, sulfamides, sulfamides-trimethoprim, tetracyclin. For each antimicrobial, the strains were classified either as susceptible or as not susceptible when intermediate or resistant. The subtypes were defined through Multiple Correspondence Analysis (MCA) and mixed classification (Berge, Atwill et al. 2003) applied on the obtained antimicrobial profiles.

## 2.1.4 Consumption

The data on food consumption used in the model come from a national survey on individual food consumption, the INCA study, conducted by the French Food Safety Agency (AFSSA) in 1999 on 3 003 representative subjects above 3 years old (Volatier 2000). We also used data

on the amount of the different meat types, eggs and sea products available for consumption, published by the French Livestock Institute and the French Interprofessional National Board for Sea and Fish-Farming Products on an annual basis. These data are converted as an annual consumption index with 1999 as reference year.

The consumption data (in tons) from the 1999 INCA study were updated with this index to obtain the 2005 consumption data. This allows having information as close as possible from the effective consumption.

## 2.2 Models

### 2.2.1 Simple model

The expected number of human cases of type i (i=1,...,I) linked to a source j (j=1,..., J) ($\lambda_{ij}$) is assumed to be proportional to the prevalence of type i in source j ($p_{ij}$) and to the exposition of the population to source j, measured by the amount of source j consumed by the general population ($M_j$). The total observed number cases of type i ($o_i$) is distributed among the sources in which type i is present, according to the relative weight :

$$\lambda_{ij} = \frac{p_{ij} \times M_j}{\sum_j p_{ij} \times M_j} \times o_i. \qquad (1)$$

Only 3% of the human Enteritidis and Typhimurium strains have been subtyped, the 97% not subtyped have been reallocated. Similar subtype distribution of cases subtyped and not subtyped has been assumed and the observed types distribution of the cases that have been subtyped within each serotype was applied to allocate the cases with unknown subtype. This reallocated number of cases was used in place of $o_i$.

The expected number of cases due to source j ($source_j$) is obtained by summing the $\lambda_{ij}$ on i. This simple approach assumes that all the sources are equivalent vehicles for the pathogen

and all the serotypes have the same capacity to induce infection, but allows a clearer inside in the functionality of the approach (Zwietering 2008).

## 2.2.2 Bayesian model

The Bayesian microbial subtyping attribution model, based on the Hald model (Hald, Vose et al. 2004), has been described in a former publication (David, informative information). As for the simple model, the number of human cases caused by different *Salmonella* types is compared with the prevalence of those types in the reservoirs, i.e. the food-animal sources, taking into account the exposition, i.e. the amount of food source consumed. Moreover, with some informative prior, the Bayesian framework allows taking the differences between sources in their capacity to vehicle the pathogen and between the types, in their capacity to cause infection, into account. The expected number of observed human cases ($o_i$) is assumed to be Poisson distributed:

$$o_i \sim \text{Poisson} \left( \Sigma_j \lambda_{ij} \right) \text{ with } \lambda_{ij} = M_j \, p_{ij} \, q_i \, a_j \qquad (2)$$

Thus, the expected number of human cases due to a given type i in a given source j ($\lambda_{ij}$) depends on the prevalence of the type i in the source j ($p_{ij}$), on the consumption of the source j in the general population ($M_j$) and on a source-dependant and a type-dependant parameters ($a_j$ and $q_i$ respectively).

As defined by Hald *et al*, the type dependant factor ($q_i$) summarizes the characteristics of the serotype (survivability, virulence, pathogenicity, …) which determine its capacity to cause an infection and the source dependant factor ($a_j$) summarizes the characteristics of the source (physical properties, preparation methods, processing procedures, …) which determine its capacity to act as a vehicle for *Salmonella* (Hald, Vose et al. 2004). The expected number of cases due to source j ($source_j$) is obtained by summing the $\lambda_{ij}$ on i.

As for the simple model, the subtypes distribution is assumed to be similar in subtyped and not subtyped cases. The cases not subtyped are reallocated within Typhimurium or

Enteritidis according to the observed subtypes proportions on a bayesian way, using Gamma distributions to reflect the uncertainty on the observed number of cases per subtype. The proportions used for the reallocation within a serotype are thus following:

$$\frac{Gamma(o_i;1)}{\sum_i Gamma(o_i;1)} \qquad (3)$$

where $o_i$ is the observed number cases of subtype i for a given serotype, Enteritidis or Typhimurium (Hald, Vose et al. 2004).

The priors were defined as we recommended (David, informative information). The type-dependant parameters relative to specific types (i.e. types only present in one food-animal source) other than Enteritidis and Typhimurium, as to avoid any interaction with the reallocation process, were fixed to the following data-based value:

$$q_i = \frac{o_i}{\sum_i o_i} \times \frac{1}{p_{ij}} \qquad (4)$$

where $o_i$ is the observed number of cases due to type i, $p_{ij}$ is the prevalence of type i in the unique source j. We thus used the percentage of human cases divided by the prevalence in the source as an indicator for factor q. The prior distributions for the other type dependant parameters and the source dependant parameters are uniform distributions with 0 as lower value and an upper value defined as to encompass all possible values for the parameter.

This parameterization requires the availability of at least J specific types not corresponding to Enteritidis and Typhimurium, which was the case for both the active and the passive datasets. It also requires the microbiological types to be heterogeneously distributed among the sources considered.

For the passive dataset, types distributions are used in place of prevalences.

# 3 Results

## 3.1 Data

### 3.1.1 Human cases

The dataset included 9 076 domestic sporadic confirmed human cases. Enteritidis and Typhimurium were the most frequent serotypes, representing respectively 35% and 39% of the cases. The subtyping resulted in 9 subtypes for Enteritidis and 9 also for Typhimurium, some of which gather only human strains or only animal strains.

### 3.1.2 Datasets

For each dataset, serotypes or subtypes which have caused only a few (or no) infections or are not observed in the animal reservoirs are grouped into the categories "others" for serotypes and "other Typhimurium" or "other Enteritidis" for the subtypes. Subsequently, the types' composition of the "other" categories is different for the human data and for each of the sources. The relative human cases are consequently considered as not attributable and will automatically be affected to the "unknown source" category in the attribution process.

We here use two datasets with different surveillance systems for the food-sources. The number of included sources and the types observed among the food-sources strains thus differ between both datasets. As a consequence, the observed serotypes common to the human cases and the food-sources are different between the two datasets which leads for the human cases as well as for each of the common food sources to differences in the composition of the categories "Others", "Other Typhimurium" and "Other Enteritidis" in the active and in the passive dataset. For the same reasons, the number of attributable cases and the types taken into account for these attributable cases differ for both datasets.

<u>Active surveillance dataset</u>

5 938 human cases are attributable to the five considered food sources, spread between 28 serotypes, 5 Typhimurium subtypes and 3 Enteritidis subtypes. The global observed prevalences in the sources were 30.8% for layers, 18.8% for pigs, 13.8% for turkeys, 8.2% for broilers and 2.4% for cattle. The types distribution among the human and in the sources is presented in figure 1a. The types are heterogeneously distributed among the sources according to Fisher exact tests.

Among the ten most frequent types within the attributable human cases, 6 are Typhimurium or Enteritidis subtypes. Overall, Enteritidis and Typhimurium represent over 75% (respectively 47.1 and 30.4%) of the human attributable cases. The human cases are mostly associated to SE-multiS (35.2%), ST1 (17.5%), SE1 (10.4%), ST4 (6.5%) and Infantis (2.8%).

The food-source cattle is characterized by a small number of types, 2 when excluding the "others" category, namely Indiana and Mbandaka, this last serotype representing 62.5% of all cattle strains. Layers are mainly associated to Infantis (13.2% of the strains), Anatum (12.6%), Bareilly (7.1%), ST5 (6.8%) and Mbandaka (6.8%). For pigs, turkeys and broilers, three serotypes each appear to be much more frequent than the others: Derby (35.9%), ST1 (19.8%) and ST5 (13.2%) for pigs ; Derby (25.3%), ST3 (13.4%) and Mbandaka (11.2%) for Turkeys ; Hadar (17.7%), Virchow (17.4%) and Montevideo (14.9%) for broilers. Thus, Derby appears to be the most frequent serotype both in the turkeys and broilers sources. Moreover, when pooling Enteritidis and Typhimurium subtypes (excluding the "other" categories), Enteritidis is mainly observed in the layers source, where it represents 7.9% of the strains. Typhimurium, if ubiquitous, is mostly associated to pigs (38.7% of the strains) and to a less extent to layers (14.3%) and to broilers (13.4%).

This dataset globally comprises 14 specific types (table 1), including one Typhimurium subtype and one Enteritidis subtype. All sources but cattle are represented among the specific types, pigs and layers being the most frequently concerned.

6 527 human cases are attributable to the 9 included sources, spread between 37 serotypes, 5 Typhimurium subtypes and 4 Enteritidis subtypes. The proportion of the types within a food-animal species are given for 617 strains for layers, 1 273 strains for broilers, 796 strains for turkeys, 213 strains for pigs, 71 strains for cattle, 104 strains for sheep, 40 strains for the sea products, 3 113 strains for ducks and 722 strains for other poultry. The types distribution among the human cases and in the sources is presented in figure 1. The types are heterogeneously distributed among the sources according to Fisher exact tests.

Among the ten most frequent types within the attributable human cases, 7 are Typhimurium or Enteritidis subtypes. Overall, Enteritidis and Typhimurium represent over 75% (respectively 47.6 and 27.7%) of the human attributable cases. The five main types are almost the same as for the active dataset, that is SE-multiS (32.1%), ST1 (15.9%), SE1 (9.4%), ST4 (5.9%) and SE3 (4.7%).

Layers, broilers, other poultry and turkeys are characterized by the diversity of serotypes and subtypes observed (29, 27, 27, 26 respectively). Layers and broilers are especially associated to Senftenberg (19.8% and 13.0% of the strains respectively) and SE-multiS (16.9% and 16.6% respectively). For turkeys, the most frequent serotype is Senftenberg (14.7%). Two serotypes emerge for other poultry: Indiana (31.0%) and ST2 (14.5%). Serotype Indiana is also the most frequent serotype in ducks (23.1%). In the passive dataset, the types associated to cattle are more diverse, 11 serotypes (vs 2 in the active dataset), the most frequent ones being ST1 (22.5%), Derby (19.7%) and Montevideo (14.1%). Sea products are mostly associated to Weltevreden (25%). At last, sheep is the source with the smallest number of types, that is 7. This source is particularly associated to the serotype S 61:k:1,5,7, which gathers 57.7% of the strains.

When pooling Enteritidis and Typhimurium subtypes (excluding the "other" categories), Enteritidis is mostly found in layers (24.3%) and broilers (16.7%), whereas Typhimurium is

rather associated to other poultry (30.7%), pigs (23.5%), cattle (22.5%) and sea products (15.0%).

This dataset comprises 15 specific types (table 1), including two Enteritidis subtypes. As for the active surveillance, all sources except cattle are represented among the specific types and pigs and layers are the most frequently concerned, in number of specific types and in number of associated human cases. Four specific types are common for both samples: SE2, Oranienburg and Havana, linked to layers and Goldcoast linked to pigs.

Thus, the included serotypes, even if globally similar, are not exactly the same in both datasets. This leads to a difference in the number of attributable cases, which is 10% higher in the passive dataset. When considering the 5 common food-animal species, the types distribution appears different between the two datasets for a given species, which was confirmed by Fisher exact tests. However, in both samples, Derby is the most frequent serotype for the pigs, which is also true for Typhimurium globally. And SE-multiS is the most frequent type in the layers whichever the source of the data is. Specific types were identified for all food sources except cattle in both datasets, though there are few common ones which concern only layers and pigs. Actually, it appears that each type of surveillance gives a different "image" of the food-sources contamination, with significantly different types repartitions for common food sources.

### 3.1.3 Consumption data

On the basis of the updated results of the INCA study, we assumed that the national consumption for 2005 was 82 301 tons for layers (eggs), 84 842 tons for broilers, 18 967 tons for turkeys, 161 971 tons for pigs, 144 504 tons for cattle, 20 232 tons for sheep, 111 764 tons for sea products, 5 309 tons for ducks and 4 580 tons for other poultry. Here appears that the sea products is a major potential source in term of consumption, though little information is available on the *Salmonella* contamination of such products.

## 3.2 Impact of the data characteristics on the attribution results using the simple model

To analyze the impact of the data characteristics on the attribution results, we used the simple model in order to beneficiate of a clear insight in its functionality.

### 3.2.1 Contamination of the sources: proportion vs prevalence

For the passive dataset, prevalences are not available, but only strains proportions per type. We here assessed the impact of using proportions instead of prevalences with the active surveillance dataset, for which both indicators were available (figure 2).

No change in the ranking of the sources was observed, but the numbers of cases attributed to each source differed. For the most important source, layers, using proportions led to a decrease of 25.7% (3 164 vs 2 350 cases respectively). Inversely, an increase in the number of attributed cases was observed for the four other food-sources, 8.2% (1 901 vs 2 057) for pigs, which is the second most important source and 84.4% for broilers (925 vs 501), 73.5% for cattle (62 vs 36) and 62.2% for turkeys (544 vs 336). For cattle, though, the number of attributed cases remained modest.

Thus, it appears possible to use proportions instead of prevalences, though, in this case the numbers of cases per source have to be interpreted with caution and the relative importance of the sources seems to be more reliable.

### 3.2.2 Increasing sources' number

The passive dataset allows studying the effect of including more sources. The model was first estimated with the five main sources, common with the active dataset, then with the nine available sources.

When including nine sources, the total number of attributable cases increased slightly, by 1.2% (figure 3). For each of the main sources, the number of attributed cases decreased, from 2.1% for layers to 12.5% for pigs, but the relative importance of those 5 sources remained the same. The cases attributed to the 4 additional sources thus correspond partly to newly attributed cases (1.2%) and partly to reattributed cases (around 7%) which were probably wrongly attributed when considering only five sources.

Thus, including additional sources allows attributing more cases and avoiding some misattributions, without changing the relative importance of the initially included sources.

Finally, comparing the number of attributed cases for the main sources between active and passive samples, using the prevalences for the active dataset (figures 2 and 3), two major differences appear which concern broilers and cattle. First, with the passive dataset, broilers is the second most important source, after layers, whereas with the active dataset broilers was at the third place with a much lower number of attributed cases. Namely, an increase of 171% is observed with the passive data (1 359 cases vs 501 cases). Second, cattle which is the less important source by far when considering the active dataset (36 cases) becomes the third source when using the passive dataset, with a quite high number of associated cases (1 170 cases).

When examining the impact of the data characteristics on the attribution results obtained with a simple model, it appeared that using prevalences or proportions had no major impact on the attribution results, as did the inclusion of additional sources. The important increase in the number of cases associated to broilers and cattle when using passive surveillance data is probably linked to the observed serotypes. Both sources present namely differences in the types repartition for the common sources. And as the representativeness is not assured for the passive data on food source contamination, interpreting results based on active surveillance data seems more convenient.

189

## 3.3 Attribution results using Bayesian model

The bayesian model is applied separately with the active and passive datasets. Results presented correspond to runs of 100 000 iterations of the Gibbs sampler with a thin of 25. Convergence diagnostics were satisfactory (Cowles and Carlin 1996; Brooks and Gelman 1998; Brooks and Roberts 1998b). From these runs, parameters estimates (posterior means, posterior variances and posterior 95% credibility intervals (95% CI) were computed from the last 50 000 iterations.

The expected numbers of cases per type and the observed numbers of cases, or the deterministically reallocated numbers of cases for Enteritidis and Typhimurium subtypes, are compared. They show a good adequacy for both datasets (Figure 4).

The total expected number of cases is, as expected, higher with the passive surveillance, 7 122 (95% CI: 6 719–7 528) compared to 5 746 (95% CI: 5 307 − 6 172) with the active surveillance dataset. For the active surveillance, the expected number of cases is in agreement with the 5 938 attributable cases, however, for the passive surveillance, it is slightly overestimated in regard to the 6 527 attributable cases.

For the 5 main sources, the numbers of expected cases per source are not significantly different when using the passive dataset, except for broilers (figure 5). It is lower for layers (2 328 vs 3 060), turkeys (336 vs 714) and pigs (1 039 vs 1 477) and higher for cattle (133 vs 69). For broilers, the number of cases is significantly higher with the passive dataset (1 410 vs 427) which was also observed with the simple model. This is probably linked to the higher frequency in this source of Typhimurium and Enteritidis subtypes in the passive dataset (27.4% globally) than in the active dataset (5.1% of the strains). Moreover, unlike when using the simple model, the number of cases attributed to cattle isn't significantly different for both datasets. This confirms the interest of the bayesian modelling and the relevance of introducing source and type dependant parameters. Actually, the source

dependant parameter relative to cattle is significantly smaller for the passive dataset (0.0039, 95% CI: 6.6E-5 − 0.018 vs 13.5, 95% CI: 0.025 − 83.6 for the active dataset), whereas, for the other common food-sources the posterior means are not significantly different (Table 2).

Except for cattle, the source dependant factors are not significantly different between the two datasets for the main sources. Even if not significantly, the best vehicle among those four sources for *Salmonella* is turkey meat, which also has the widest 95% CI. The other sources are very close. No firm conclusion can be drawn for cattle, which source dependant factor is the highest for the active dataset and the smallest for the passive dataset, with very wide 95% CI. Finally, in the passive dataset, the highest source dependant factors are relative to ducks (1.37) and other poultry (1.06), the difference with the other sources being significant. The better vehicle for *Salmonella* thus appears to be birds, which are also especially linked to Enteritidis, the most frequent serotype among the attributable human cases.

Concerning the type-dependant parameters, the posterior means obtained with the passive dataset are systematically lower than the ones obtained with the active dataset, except for Stourbridge, mostly significantly. Another global feature is that the 95% CI are less wide for Enteritidis and Typhimurium subtypes with the passive dataset. When considering the 10 highest type dependant parameters for each dataset (table 3), 6 types are common: SE1, SE-multiS, Stourbridge, Havana, Oranienburg and Napoli, but only the parameter relative to Stourbridge has the same value in both dataset, which is a fixed value in the active dataset (5.40) and an estimated value in the passive dataset (5.41 95% CI: 2.42 − 10.72). For Napoli, if the values are different, the rank is the same in both cases: 10. It can also be observed that 3 highly ranked parameters in the active dataset (corresponding to Heidelberg, ST2 and SE2) are low ranked in the passive dataset, respectively 26th, 33rd and

19th rank, instead of 1st, 3rd and 8th rank. For what concerns the passive dataset, 4 types in this top ten are not observed in the active dataset, all the parameters relative to the other types, are all also in the top ten in the active dataset. At last, when considering Enteritidis and Typhimurium relative parameters, it can be seen that all Enteritidis subtypes but one, are in the top ten for each dataset, which underlines the infectious capacity of this serotype, and the only Typhimurium subtype appearing here is ST2 in the active dataset, which gathers multi-resistant strains resistant to nalidixic acid.

To summarize the results, with the active surveillance dataset, the major source of infection is layers (eggs) to which 53.3% of the expected cases are attributed (95% CI: 46.0-60.0), followed by pigs, to which 25.7% (95% CI: 20.7-31.1) of the cases are attributed. The third place is taken by broilers and turkeys which results are not significantly different, respectively 7.4% and 12.4% (95% CI: 5.6-9.5 and 6.9-19.2). Cattle appears to be a minor source, which comprised only 1.2% of the cases (95% CI: 0.2-1.7). The cases which origin remains unknown are as numerous as the expected number of cases attributed to layers.

Using the passive dataset leads to comparable results concerning layers, pigs, turkeys and cattle, though for broilers, the number of attributed cases is significantly higher (1 410, 95% CI 1 026-1 821), so that this source is comparable to pigs (1 039, 95% CI: 785-1 310) in importance. The ranking of the sources so would be layers as the most important source of infection, broilers and pigs as second importance sources, all the other sources (turkeys, cattle and the four additional) following, without distinction as the expected numbers of attributed cases are near from each other and the 95% CI quite wide. As for the results obtained with the active dataset, the cases of unknown origin are as numerous as the number of cases attributed to layers.

Concerning a and q parameters, it is difficult to draw any firm conclusion on the comparative infectious capacities of the different serotypes as well as on the relative capacity of the sources to vehicle the pathogen. Namely, when some parameters are higher than the others,

their 95% CI are wide. Moreover, the results even in term of ranking diverge between both datasets. Though, it can be extrapolated from the few convergent results, that birds appear to be a good vehicle for *Salmonella* and that Enteritidis has a high capacity of infection, whichever the subtype is, as do some serotypes such as Stourbridge, Havana, Oranienburg and Napoli.

# 4 Discussion

The application of the microbial subtyping approach as conceived by Hald *et al* requires intensive monitoring data on human cases and all relevant food-sources, moreover the contamination of sources should be measured in prevalence and discriminatory typing methods, beyond serotyping, should be applied (Hald, Vose et al. 2004). Producing representative prevalence data at the farm and at the abattoir levels is essential. It necessitates to actively monitor the main food-sources, which is costly and not always developed for this early point of the food-chain (Batz, Doyle et al. 2005; Kirk, McKay et al. 2008; Sofos 2008). Thus, passively acquired surveillance data can be the only available alternative, even if their representativeness is not assured. In France, thanks to the European regulatory frame and to long lasting lab-based surveillance systems, all the required data were available and the typing methods were harmonized for all datasets, as a consequence of the ongoing cooperations for outbreaks investigations. Moreover, we disposed for the sources both of representative prevalence data and of passive surveillance data, which allowed us assessing the relative advantages of using such data in the bayesian attribution approach in terms of results quality and of potential sources coverage.

A large scope collaboration, which is a key point for such projects (Batz, Doyle et al. 2005; ICMSF 2006), allowed centralizing all the required data, which were fragmented between several actors. This led to the build an active and a passive dataset. The heterogeneity in the repartition of the types among the sources was met, as did the necessary number of specific types for the parameterization of the model. Though, we had to deal with some insufficiencies in the data retrieved. The human database is produced by a lab-based surveillance network, thus neither exhaustive, nor definitely representative. Moreover, the information on travel was not always accurate. Though, as it relies on a stable and nationwide network of public and private laboratories, which representativeness is estimated as good, it constitutes a reasonable basis for an attribution work. Another problem is the lack of time consistency of the actively acquired food-source data with the human cases. This raises concern about a potential bias on the results obtained with the active dataset. Though, this will be solved for the coming years, as the main sources (layers, broilers, turkeys and pigs) will be regularly monitored as required by the European Zoonosis Directive which is actually enforced (David, Danan et al. submitted). Finally, the subtyping of human Enteritidis and Typhimurium strains concerned only 3% of the cases, the other cases being reallocated in the model, which led to wide credibility intervals for the type-dependant parameters and contributed to the uncertainty on the attribution results. Moreover, assuming similar distribution of cases subtyped and not subtyped seems in these conditions questionable. Testing more human strains for antimicrobial resistance would certainly enhance the predictions of the model. We thus disposed of a complete dataset, based on harmonized typing methodologies, but, because of some imperfections, and especially of the absence of any data on the imported products and a lack of time consistency of the active data, the attribution results presented, if valuable to study the methodological points (passive vs active data), have to be taken with caution regarding their representativeness and should be enhanced in future studies.

The main point studied was the comparison of the results obtained with passive and active data regarding the source contamination in the model. First of all, the image in term of types distribution given by the both surveillance systems for the common sources were significantly different. Moreover, the data characteristics were different in term of contamination indicator and of number of included sources. Using the simple model, these characteristics had little impact on the attribution results when considered singly, but when comparing the results obtained for the differences altogether, they were quite different (especially for broilers and cattle). The bayesian model, on the contrary, seems to be robust to this heterogeneity, as the results are in accordance for both datasets, except for broilers. Thus, passive data appear to be a potential complement for active datasets. Namely, the active and systematic surveillance of food sources is costly and thus can only encompass a limited number of sources. Including additional sources through the passive surveillance has several advantages. As one of the assumptions of the model is that all the attributable cases are linked to the included sources, including as many potential sources as possible allows avoiding misattributions. Moreover, as the contamination of the sources at the farm level evolves according to public health interventions, enforcement of regulations and changes in hygiene practices (Rostagno, Hurd et al. 2005; Denagamage, O'Connor et al. 2007), and as the exposition of the population to the sources also changes (Combris 1997; Allard 2002; Desenclos, Vaillant et al. 2002) the model can help detecting the emergence as public health concerns, of sources not institutionally monitored but which should be. Finally, it is cost-effective, which is an important point in public health tools design (Havelaar, Bräunig et al. 2007). Thus, when applied with both active and passive data, the model could be a useful tool to contribute to the adjustment of the institutional active surveillance, besides its potential role for the conception and the evaluation of public health interventions.

If only passive surveillance data are available for the sources contamination, reasonably valuable results are obtained. Of course in this case, conclusions will have to be made with caution, taking the quality of the data into account.

The results obtained on the relative importance of the sources, when considering the active dataset complemented by the passive dataset for the 4 additional sources, are in accordance with the Danish results (Hald, Vose et al. 2004; Havelaar, Bräunig et al. 2007): layers being the most important source, followed by pigs and broilers, turkeys, ducks and cattle which are not significantly different. Broilers is the only source for which the expected number of cases is significantly different between both datasets. This could be due to the higher frequency of Typhimurium and Enteritidis among the strains in the passive dataset, the source-dependant factor being similar between both datasets. Such a difference in the results obtained for *Salmonella* with active and passive data especially for broilers (compared to pigs) has been recently exposed in a different area (antimicrobial resistance) (Mather, Haydon et al. 2009). Thus, it would be interesting to explore the reasons of this difference, specific to broilers.

For each configuration, the number of cases with unknown origin was as high as the number of expected cases attributed to the main source, i.e. layers. The major part of these "unknown" cases belonged to Enteritidis and Typhimurium "other" categories, corresponding to subtypes not observed in the considered animal sources. An explanation for this could be either a bias due to the small proportion of strains subtyped among the human cases or a detection problem of rare types in the sources. Though, in this last case, the proportion of unknown cases should be much lower with the passive dataset, as the passive surveillance allowed gathering more strains for each common animal species and to include more sources and thus potentially better detect rare types. Another probable reason for this high frequency of cases with unknown reservoir, is to be found in the potential sources not considered in this analysis. We namely only considered the foodborne transmission, thereby excluding sources such as pets, and exotic pets which are known sources of human Salmonellosis (Woodward, Khakhria et al. 1997; Bellido Blasco, Gonzalez Cano et al. 1998; Mermin, Hutwagner et al. 2004; De Jong, Andersson et al. 2005; Finley, Reid-Smith et al. 2006; Marcus, Varma et al. 2007) or even man to man transmission (Todd, Greig et al. 2008). Moreover among the potential food sources, we focused on some animal food-

sources, which are certainly the most frequent sources of human Salmonellosis but not the only ones. Plants (Brandl 2006; Elviss, Little et al. 2009) and other non animal products (Kirk, McKay et al. 2008) have been implicated as a source for Salmonellosis. At last, imported products, though representing from 15% (for eggs) to 60% (Lamb) of the national consumption according to the food-animals products, could not be taken into account in this analysis but also are potential sources for these unknown cases.

Direct comparison between the source dependant factors for the two datasets was difficult. Indeed, the contamination indicator is prevalence in one case and proportion in the other case and as a consequence, the source dependant parameter doesn't measure exactly the same thing, which is also true for the type dependant factor. Actually, the type dependant factors were almost systematically different when using one or the other dataset, even when the ranking was coherent. The estimated source dependant factors though, were in good agreement between both datasets, except for cattle. Moreover, the 95% CI were mostly too wide to obtain significant differences.

Hald et al underlined that these factors were only multiplication factors allowing arriving at the most probable solution (Hald, Vose et al. 2004). This was actually the case in our study, especially when considering cattle, for which the attribution results diverged using the simple model, but were in accordance when introducing the a and q parameters. Though, it was difficult to draw any firm conclusion on the obtained posterior distributions, especially for the type dependant parameter. Actually, those parameters are conceived as black boxes, they allow taking into account differences between types and between sources, which is of primarily importance (Blaser and Newman 1982; D'Aoust 1989; Sarwari, Magder et al. 2001; Coleman, Marks et al. 2004; Bollaerts, Aerts et al. 2008; Jones, Ingram et al. 2008), but their nature is not defined. It thus would be interesting to give a definition to these parameters, which would allow to use exogenous information, such as dose-response relationship, infective dose, pathogenicity (Bollaerts, Aerts et al. 2008; Jones, Ingram et al. 2008),…, to validate informative prior distributions.

## Conclusion

We here addressed the problem of the data quality, taking advantage of the existence in France of an active and a passive surveillance system for what concerns the contamination of the sources. Our results indicate that it is possible to obtain reasonably reliable attribution estimates when using passively acquired data and even that those data can be an interesting and cost-effective complement to the active and representative prevalence data. Though, the results for the additional sources will remain indicative and conclusions should be drawn on the ranking rather than on the absolute number of cases. Moreover, the sources and type dependant parameters can hardly be interpreted, given the differences between both datasets results. For France, the improvement of the approach should be obtained by an improvement of the human and animal datasets, for example by a better collection of data about travel history of the human cases, the sub-typing a larger subset of Enteritidis and Typhimurium human strains and the collection of data about prevalence of salmonella in main imported animal sources. These recommendations applied to further studies would lead to a decrease of uncertainty around estimates. The Bayesian tool could also be improved by a research work about the nature and dimension of the type and source parameters which are of primarily importance in the approach. Finally, the parallel use of active and passive data could contribute to a risk-based orientation of the active surveillance policy.

# References

Allard, D. G. (2002). "The 'farm to plate' approach to food safety - Everyone's business." Canadian Journal of Infectious Diseases **13**(3): 185-190.

Batz, M. B., M. P. Doyle, et al. (2005). "Attributing illness to food." Emerging Infectious Diseases **11**(7): 993-999.

Bellido Blasco, J. B., J. M. Gonza?lez Cano, et al. (1998). "Factors associated with sporadic cases of salmonellosis in 1- to 7-year-old children. Study of cases and controls." Factores asociados con casos espora?dicos de salmonelosis en nin?os de 1 a 7 an?os. Estudio de casos y controles.

Gaceta sanitaria / S.E.S.P.A.S **12**(3): 118-125.

Berge, A. C. B., E. R. Atwill, et al. (2003). "Assessing antibiotic resistance in fecal Escherichia coli in young calves using cluster analysis techniques." Preventive Veterinary Medicine **61**(2): 91-102.

Blaser, M. J. and L. S. Newman (1982). "A review of human salmonellosis: I. Infective dose." Reviews of Infectious Diseases **4**(6): 1096-1106.

Bollaerts, K., M. Aerts, et al. (2008). "Human salmonellosis: estimation of dose-illness from outbreak data." Risk Anal **28**(2): 427-40.

Brandl, M. T. (2006). "Fitness of human enteric pathogens on plants and implications for food safety." Annu Rev Phytopathol **44**: 367-92.

Brooks, S. P. and A. Gelman (1998). "General methods for monitoring convergence of iterative simulations." Journal of Computational and Graphical Statistics **7**(4): 434-455.

Brooks, S. P. and G. O. Roberts (1998b). "Convergence assessment techniques for Markov chain Monte Carlo." Statistics and Computing **8**(4): 319-335.

Coleman, M. E., H. M. Marks, et al. (2004). "Discerning strain effects in microbial dose-response data." J Toxicol Environ Health A **67**(8-10): 667-85.

Combris, P. (1997). "The consumption of animal products in France : current tendencies and future perspectives." Productions Animales **10**(4): 267-274.

Cowles, M. K. and B. P. Carlin (1996). "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review." Journal of the American Statistical Association **91**(434): 883-904.

D'Aoust, J. (1989). Salmonella. Foodborne Bacterial Pathogens. M. P. Doyle. New York, Marcel Dekker: 327-445.

David, J. M., C. Danan, et al. (submitted). "Structure of the French farm to table surveillance system for *Salmonella*."

De Jong, B., Y. Andersson, et al. (2005). "Effect of regulation and education on reptile-associated salmonellosis." Emerging Infectious Diseases **11**(3): 398-403.

Denagamage, T. N., A. M. O'Connor, et al. (2007). "Efficacy of vaccination to reduce Salmonella prevalence in live and slaughtered swine: A systematic review of literature from 1979 to 2007." Foodborne Pathogens and Disease **4**(4): 539-549.

Desenclos, J. C., V. Vaillant, et al. (2002). "[Food-borne infections: do we need further data collection for public health action and research]." Rev Epidemiol Sante Publique **50**(1): 67-79.

Edel, W. (1994). "Salmonella enteritidis eradication programme in poultry breeder flocks in The Netherlands." International Journal of Food Microbiology **21**(1-2): 171-178.

Elviss, N. C., C. L. Little, et al. (2009). "Microbiological study of fresh herbs from retail premises uncovers an international outbreak of salmonellosis." Int J Food Microbiol.

Finley, R., R. Reid-Smith, et al. (2006). "Human health implications of Salmonella-contaminated natural pet treats and raw pet food." Clin Infect Dis **42**(5): 686-91.

Hald, T., D. M. A. Lo Fo Wong, et al. (2007). "The attribution of human infections with antimicrobial resistant Salmonella bacteria in Denmark to sources of animal origin." Foodborne Pathogens and Disease **4**(3): 313-326.

Hald, T., D. Vose, et al. (2004). "A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis." Risk Anal **24**(1): 255-69.

Havelaar, A. H., J. Bräunig, et al. (2007). "Towards an integrated approach in supporting microbiological food safety decisions." Zoonoses and Public Health **54**(3-4): 103-117.

ICMSF, I. C. o. M. S. f. F. (2006). "Use of epidemiologic data to measure the impact of food safety control programs." Food Control **17**(10): 825-837.

InVS (2004). Morbidité et mortalité dues aux maladies infectieuses d'origine alimentaire en France: 192.

Jones, T. F., L. A. Ingram, et al. (2008). "Salmonellosis outcomes differ substantially by serotype." J Infect Dis **198**(1): 109-14.

Kirk, M. D., I. McKay, et al. (2008). "Food safety: foodborne disease in Australia: the OzFoodNet experience." Clin Infect Dis **47**(3): 392-400.

Marcus, R., J. K. Varma, et al. (2007). "Re-assessment of risk factors for sporadic Salmonella serotype Enteritidis infections: A case-control study in five FoodNet Sites, 2002-2003." Epidemiology and Infection **135**(1): 84-92.

Mather, A. E., D. T. Haydon, et al. (2009). Comparing active and passive surveillance of antimicrobial reistance: Are we getting the whole story? 12th International Symposium on Veterinary Epidemiology and Economics, Durban.

Mermin, J., L. Hutwagner, et al. (2004). "Reptiles, amphibians, and human Salmonella infection: A population-based, case-control study." Clinical Infectious Diseases **38**(SUPPL. 3).

Mullner, P., G. Jones, et al. (2009). "Source attribution of food-borne zoonoses in New Zealand: A modified hald model." Risk Analysis **29**(7): 970-984.

Pires, S. M., E. G. Evers, et al. (2009). "Attributing the human disease burden of foodborne infections to specific sources." Foodborne Pathog Dis **6**(4): 417-24.

Poirier, E., L. Watier, et al. (2008). "Evaluation of the impact on human salmonellosis of control measures targeted to Salmonella Enteritidis and Typhimurium in poultry breeding using time-series analysis and intervention models in France." Epidemiol Infect **136**(9): 1217-24.

Rostagno, M. H., H. S. Hurd, et al. (2005). "Resting pigs on transport trailers as an intervention strategy to reduce Salmonella enterica prevalence at slaughter." Journal of Food Protection **68**(8): 1720-1723.

Sarwari, A. R., L. S. Magder, et al. (2001). "Serotype distribution of Salmonella isolates from food animals after slaughter differs from that of isolates found in humans." Journal of Infectious Diseases **183**(8): 1295-1299.

Sofos, J. N. (2008). "Challenges to meat safety in the 21st century." Meat Science **78**(1-2): 3-13.

Todd, E. C., J. D. Greig, et al. (2008). "Outbreaks where food workers have been implicated in the spread of foodborne disease. Part 5. Sources of contamination and pathogen excretion from infected persons." J Food Prot **71**(12): 2582-95.

Volatier, J. L. (2000). Enquête nationale sur les consommations alimentaires. Tec&Doc, CREDOC-AFSSA-DGAL.

Wegener, H. C., T. Hald, et al. (2003). "Salmonella control programs in Denmark." Emerging Infectious Diseases **9**(7): 774-780.

Woodward, D. L., R. Khakhria, et al. (1997). "Human salmonellosis associated with exotic pets." Journal of Clinical Microbiology **35**(11): 2786-2790.

Zwietering, M. H. (2008). "Quantitative risk assessment: Is more complex always better? Simple is not stupid and complex is not always more correct." Int J Food Microbiol.
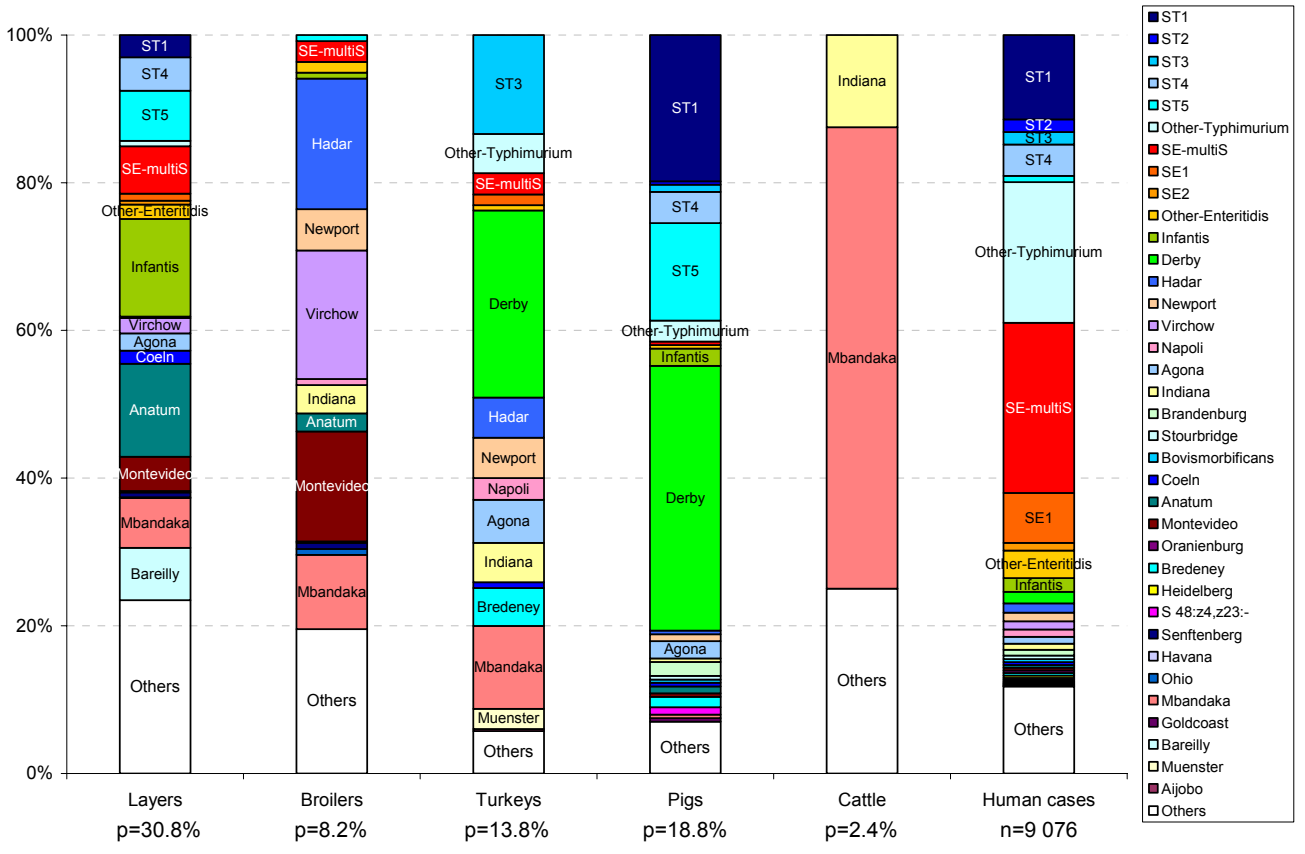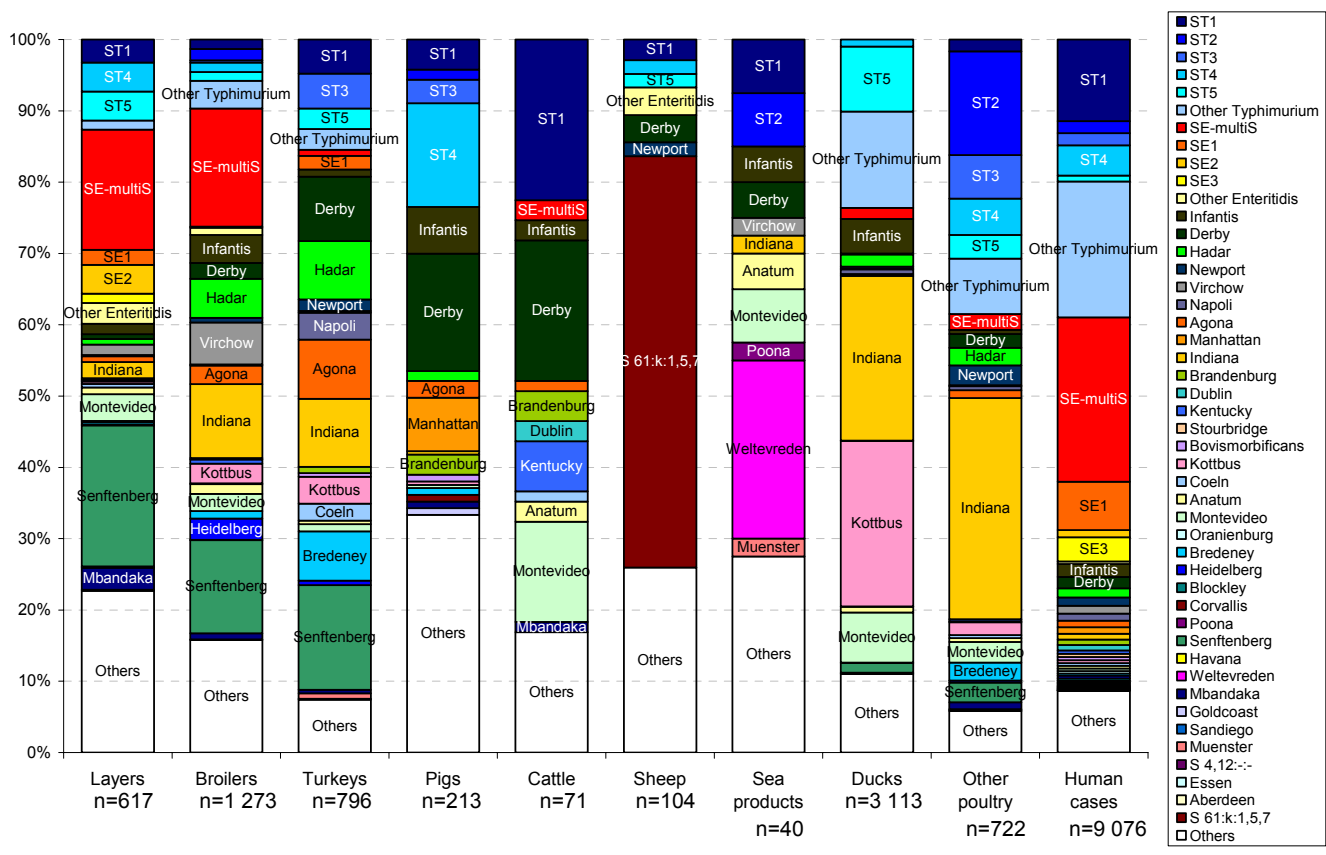
Figure 1: type distribution in the human cases and the sources according to active and passive surveillance

p: overall prevalence in the source

| Active surveillance | | | | Passive surveillance | | | |
|---|---|---|---|---|---|---|---|
| Type | Source | Prevalence (%) | Number of cases | Type | Source | Proportion (%) | Number of cases |
| Heidelberg | Broilers | 0.02 | 30 | Blockley | Broilers | 0.08 | 29 |
| Ohio | Broilers | 0.07 | 12 | SE3 | Layers | 1.30 | 308 |
| SE2 | Layers | 0.20 | 92* | SE2 | Layers | 4.05 | 92* |
| Oranienburg | Layers | 0.09 | 31 | Oranienburg | Layers | 0.16 | 31 |
| Havana | Layers | 0.04 | 16 | Havana | Layers | 0.16 | 16 |
| Bareilly | Layers | 2.95 | 8 | Manhattan | Pigs | 7.51 | 81 |
| ST2 | Pigs | 0.09 | 154* | Corvallis | Pigs | 0.94 | 24 |
| Goldcoast | Pigs | 0.09 | 9 | Goldcoast | Pigs | 0.94 | 9 |
| Stourbridge | Pigs | 0.09 | 42 | S 4,12:-:- | Other poultry | 0.14 | 4 |
| Bovismorbificans | Pigs | 0.09 | 40 | Aberdeen | Other poultry | 0.14 | 2 |
| S 48:z4,z23:- | Pigs | 0.18 | 24 | Poona | Sea products | 2.50 | 21 |
| Brandenburg | Pigs | 0.34 | 71 | Weltevreden | Sea products | 25.0 | 9 |
| Muenster | Turkeys | 0.38 | 7 | Essen | Turkeys | 0.13 | 2 |
| Aijobo | Turkeys | 0.04 | 2 | Sandiego | Ducks | 0.06 | 8 |
| | | | | S 61:k:1,5,7 | Sheep | 57.7 | 1 |

Table 1: Characteristics of the specific types
* : reallocated number of cases

| | Active dataset | | | Passive dataset | | |
|---|---|---|---|---|---|---|
| | Mean | 95CI | | mean | 95CI | |
| Layers | 0.094 | 0.070 | 0.12 | 0.097 | 0.070 | 0.13 |
| Broilers | 0.092 | 0.066 | 0.12 | 0.090 | 0.059 | 0.13 |
| Turkeys | 0.32 | 0.14 | 0.59 | 0.11 | 0.02 | 0.31 |
| Pigs | 0.054 | 0.046 | 0.062 | 0.055 | 0.045 | 0.065 |
| Cattle | 13.5 | 0.025 | 83.64 | 0.0039 | 6.57E-05 | 0.018 |
| Sheep | - | - | - | 0.54 | 0.056 | 1.62 |
| Sea products | - | - | - | 0.076 | 0.050 | 0.11 |
| Ducks | - | - | - | 1.37 | 0.59 | 2.51 |
| Other poultry | - | - | - | 1.06 | 0.38 | 2.16 |

Table 2 : Posterior means and 95% CI of the source-dependant parameters

|  | Active dataset | | | Passive dataset | | |
|---|---|---|---|---|---|---|
|  | mean | 95CI | Ranking | mean | 95CI | Ranking |
| Heidelberg | 19.24 | - | - | 1 | | 26 |
| ST2 | 14.01 | 0.65 | 36.75 | 3 | | 33 |
| Bovismorbificans | 5.14 | - | - | 6 | | 13 |
| SE2 | 4.26 | 0.17 | 12.82 | 8 | | 19 |
| SE1 | 14.16 | 8.45 | 21.67 | 2 | 2.88 | 1.61 | 4.56 | 5 |
| SE-multiS | 8.18 | 6.17 | 10.56 | 4 | 0.73 | 0.55 | 0.94 | 9 |
| Stourbridge | 5.40 | - | - | 5 | 5.41 | 2.42 | 10.72 | 1 |
| Havana | 4.41 | - | - | 7 | 1.09 | - | - | 8 |
| Oranienburg | 3.89 | - | - | 9 | 2.11 | - | - | 6 |
| Napoli | 3.23 | 1.65 | 5.80 | 10 | 0.56 | 0.26 | 0.98 | 10 |
| Blockley | | | | NO | 4.07 | - | - | 2 |
| Dublin | | | | NO | 3.19 | 0.77 | 6.26 | 3 |
| SE3 | | | | NO | 3.03 | 1.45 | 5.41 | 4 |
| Sandiego | | | | NO | 1.37 | - | - | 7 |

Table 3: Top 10 of the estimated type dependant parameters for each dataset
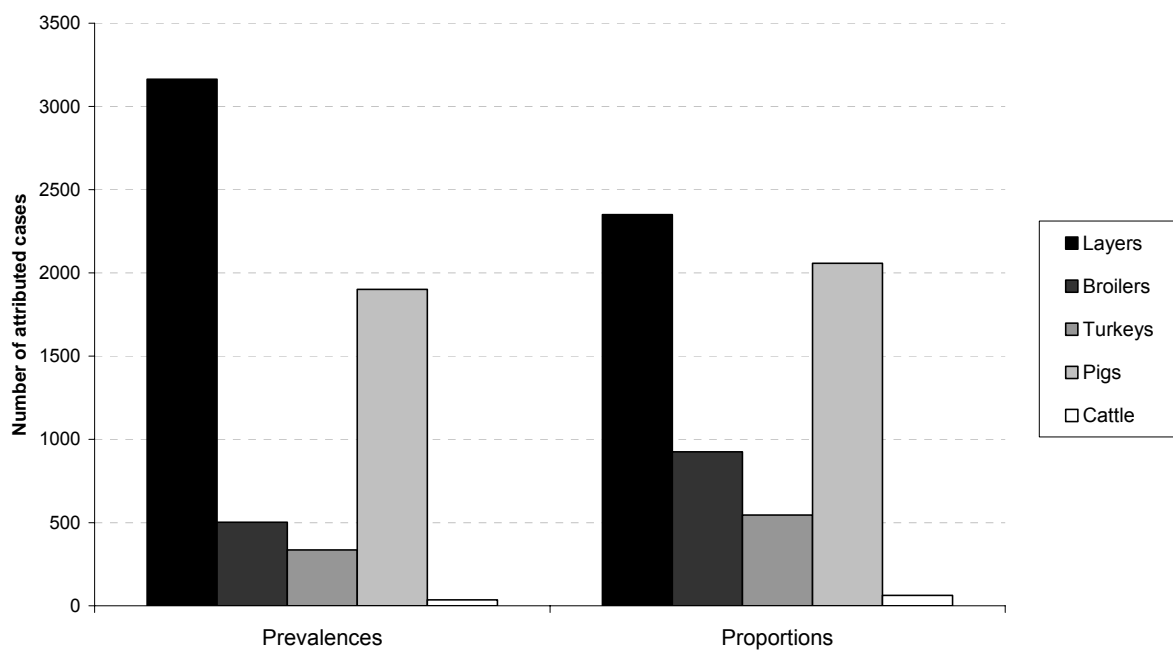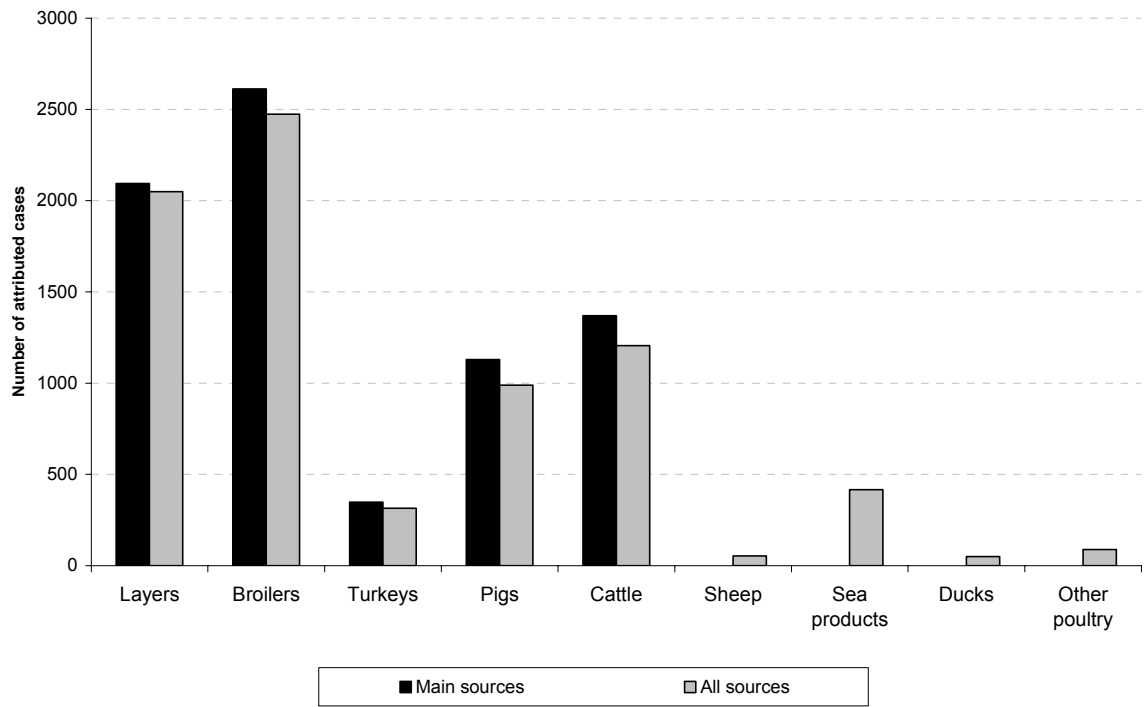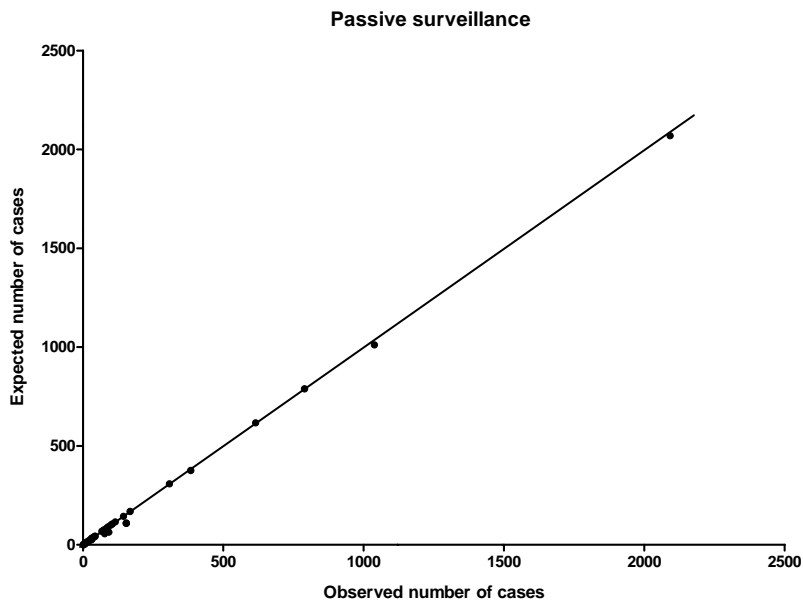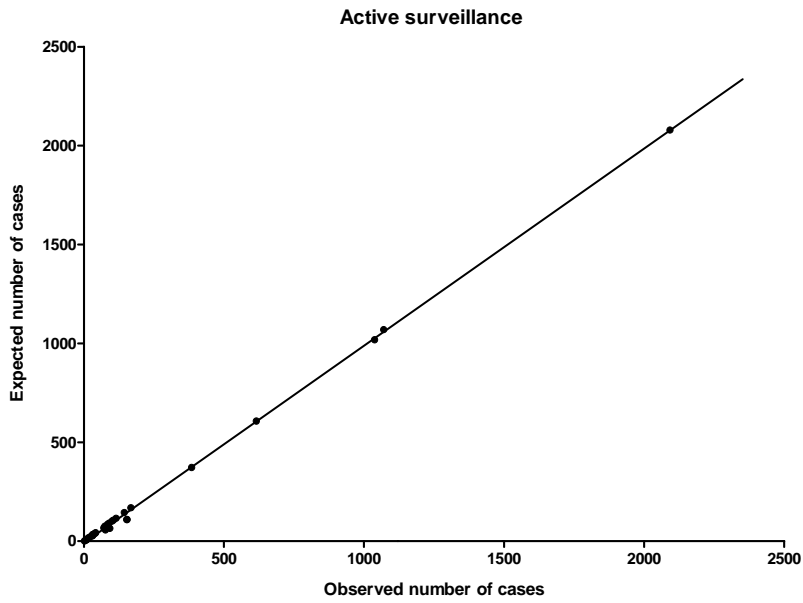NO: not observed in the dataset



Figure 2: Impact of the indicator on the attribution results, simple model, active dataset

Figure 3: Impact of the number of sources included in the model on the attribution results, simple model, passive dataset

**Active surveillance**



**Passive surveillance**



<u>Figure 4</u>: Plot of the observed and expected cases for the individual types for the active and the passive surveillance

For Enteritidis and Typhimurium subtypes, the deterministically reallocated numbers of cases are used in place of the observed numbers of cases
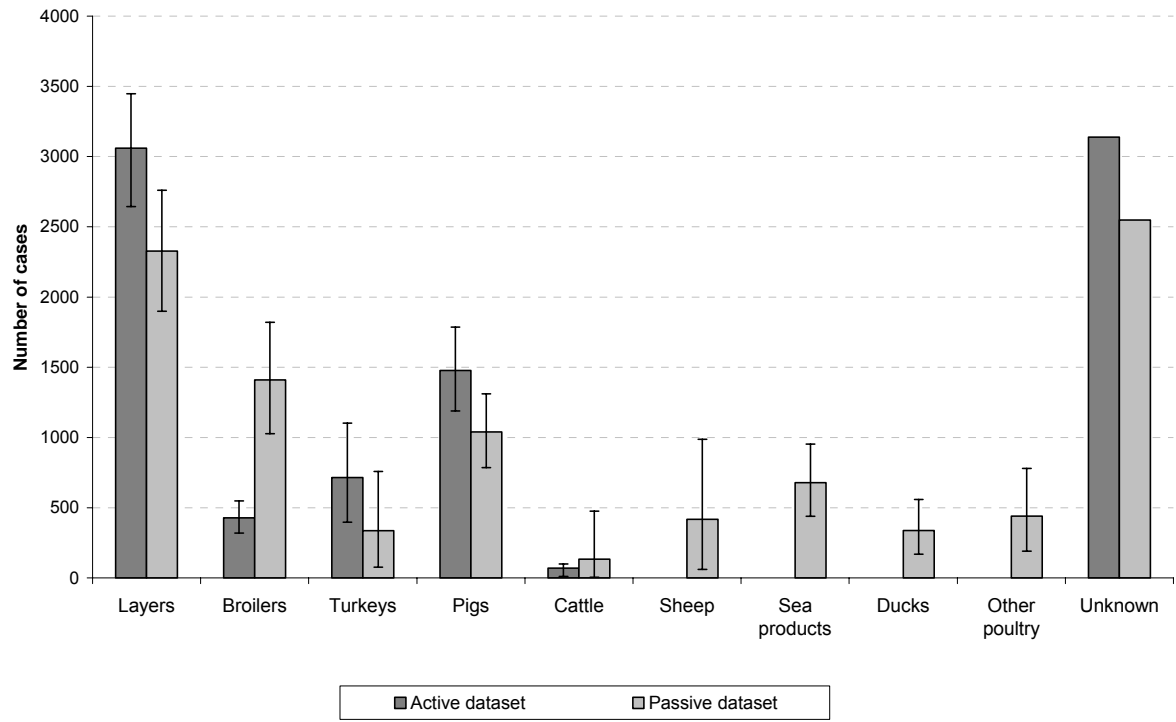
Figure 5: Posterior means and 95% CI of attributed cases per source according to surveillance mode